

Introduction aux matrices aléatoires

Djalil Chafaï*

Printemps 2013

Version du mardi 19 avril 2013

Résumé

En concevant les mathématiques comme un graphe, où chaque sommet est un domaine, la théorie des probabilités et l’algèbre linéaire figurent parmi les sommets les plus connectés aux autres. Or leur réunion constitue le cœur de la théorie des matrices aléatoires. Cela explique peut-être la richesse exceptionnelle de cette théorie très actuelle. Les aspects non-linéaires de l’algèbre linéaire y jouent un rôle profond et fascinant. Ces notes en présentent quelques aspects.

Table des matières

1 Introduction	1
2 Esquisses de preuves du théorème de Wigner	8
2.1 Méthode des moments (combinatoire)	9
2.2 Méthode de la résolvante (analyse)	11
2.3 Méthode de minimisation d’énergie pour GUE	12
2.4 Méthode des polynômes orthogonaux pour GUE	13
3 Au delà du théorème de Wigner	14
3.1 Comportement au bord et universalité locale	15
3.2 Probabilités libres de Voiculescu	16
3.3 Quelques mots sur le théorème de Marchenko-Pastur	18
3.4 Quelques mots sur le théorème de Girko	21
A Quelques outils	22
Bibliographie	25

1 Introduction

Les sources historiques de la théorie des matrices aléatoires sont multiples. Parmi les plus connues, on trouve les travaux du statisticien Wishart dans les années 1920-1930 lors de la naissance de la statistique mathématique, les travaux de Goldstine et von Neumann dans les années 1940-1950 lors de la naissance de l’informatique et de l’analyse numérique, les travaux de Wigner, Mehta, Gaudin, et Dyson dans les années 1950-1960 en physique nucléaire, à la jonction entre mécanique quantique et mécanique statistique. Il y a ensuite les travaux de Marchenko, Pastur, Girko, Bai,

*Université Paris-Est Marne-la-Vallée, Institut Universitaire de France, École Polytechnique.

et Silverstein dans les années 1960-1980, puis une multitude de travaux plus modernes, balayant un spectre étonnamment vaste incluant opérateurs de Schrödinger, physique statistique, systèmes intégrables, systèmes de particules, algèbre d'opérateurs et théorie des probabilités libres (Voiculescu), topologie et combinatoire, combinatoire additive et graphes aléatoires, polynômes orthogonaux, équations différentielles de Painlevé, problèmes de Riemann-Hilbert, analyse harmonique, théorie des nombres et hypothèse de Riemann, analyse géométrique de grande dimension, analyse numérique, classification statistique, théorie du signal et analyse d'images, etc.

La motivation originelle de Wishart était d'étudier la matrice de covariance empirique des échantillons gaussiens multivariés. Cette matrice de covariance empirique est une matrice hermitienne aléatoire semi-définie positive, appelée aujourd'hui matrice de Wishart, du même type que celle qui apparaît dans le théorème 1.2 de Marchenko-Pastur. Ces aspects, trop brièvement abordés dans la section 3.3, sont re-devenus à la mode avec l'avènement de l'informatique et des masses de données de grande dimension (big-data). La motivation originelle de Goldstine et von Neumann était de comprendre le comportement des algorithmes classiques de l'analyse numérique matricielle, comme ceux de résolution de systèmes linéaires, lorsque les données du problème sont entachées d'erreurs. Cela conduit tout naturellement à s'intéresser au conditionnement¹ de matrices aléatoires. Les problèmes qu'ils ont formulés à l'époque n'ont été résolus en toute généralité que dans les années récentes. La motivation originelle de Wigner était d'expliquer la répartition des niveaux d'énergie dans les noyaux atomiques. En mécanique quantique, les niveaux d'énergie sont liés à des spectres d'opérateurs hermitiens. L'idée de Wigner a consisté à s'inspirer de la physique statistique de Boltzmann-Gibbs, ce qui conduit à rendre ces opérateurs aléatoires, en utilisant par exemple des lois gaussiennes. La distribution des écarts entre les valeurs propres des matrices hermitiennes aléatoires obtenues de la sorte coïncide assez bien avec le phénomène observé ! L'approche pionnière de Wigner a été reprise et considérablement développée en particulier par Dyson et Mehta. Par ailleurs, à la même époque que Wigner, Philip Warren Anderson a prédit une transition entre localisation et délocalisation pour les électrons dans un semi-conducteur, phénomène qui se traduit mathématiquement par une localisation/délocalisation des vecteurs propres d'opérateurs de Schrödinger aléatoires. La compréhension mathématique de ce phénomène n'est toujours pas achevée à l'heure actuelle. En dimension 1, ces opérateurs de Schrödinger aléatoires sont typiquement des matrices hermitiennes à bande, aléatoires. En télécommunications, la transmission d'information entre un groupe d'antennes émettrices et un groupe d'antennes réceptrices est modélisable par une matrice aléatoire. La capacité du canal de communication associée s'avère être une statistique linéaire du spectre de cette matrice, ouvrant la voie à toute une gamme de questions mathématiques et numériques. Ces quelques exemples historiques illustrent la grande diversité des situations conduisant à des matrices aléatoires.

En anglais, on pourra consulter l'ouvrage encyclopédique [ABF], qui fournit un panorama récent, peut-être plus accessible que les livres [De, HP, M, ER, G1, DG, BS, AGZ, F, PS, HJ, T1, CD, EK]. En français, on pourra consulter par exemple les textes de vulgarisation [Bi1, Bi2, P] sur des sujets précis, et même s'aventurer dans l'article de synthèse [G2] qui présente des avancées relativement récentes sur l'un des fronts actuels de la recherche. Signalons enfin que plusieurs blogs Internet proposent des

1. Si $A \in \mathcal{M}_n(\mathbb{C})$, son conditionnement est $\kappa(A) := \lambda_{\max}(\sqrt{AA^*})/\lambda_{\min}(\sqrt{AA^*}) = \|A\|_{2 \rightarrow 2}/\|A^{-1}\|_{2 \rightarrow 2}$.

billets de vulgarisation autour de la théorie des matrices aléatoires, comme le blog de Terence Tao par exemple (en anglais) à l'adresse <http://terrytao.wordpress.com/>.

Ces notes constituent un point de départ, une invitation à la curiosité et à l'exploration. Nous débutons par des expériences numériques en langage Scilab dont le code ainsi que la sortie graphique sont donnés dans les figures 1-2-3. Pour une matrice $M \in \mathcal{M}_n(\mathbb{C})$, on note $M^* := \bar{M}^\top$, et $\lambda_1(M), \dots, \lambda_n(M)$ les valeurs propres de M , c'est-à-dire les racines dans \mathbb{C} de son polynôme caractéristique. La manière de les numéroter importe peu à ce stade car la plupart des quantités que nous allons considérer sont des fonctions symétriques des valeurs propres. Pour tout $n \geq 1$ fixé, soit M une matrice aléatoire $n \times n$ dont les coefficients sont des variables aléatoires sur \mathbb{C} (c'est-à-dire des vecteurs aléatoires de \mathbb{R}^2) i.i.d.² de moyenne

$$m := \mathbb{E}(M_{ij}) = \mathbb{E}(\Re M_{ij}) + \sqrt{-1}\mathbb{E}(\Im M_{ij})$$

et de variance

$$\sigma^2 := \mathbb{E}(|M_{ij} - m|^2) = \mathbb{E}(|M_{ij}|^2) - |m|^2 = \text{Var}(\Re M_{ij}) + \text{Var}(\Im M_{ij}) > 0.$$

Ce modèle tolère tout à fait que M soit réelle. La matrice M de nos expériences numériques 1-2-3 a des coefficients i.i.d. réels de loi de Rademacher ou de Gauss. Rappelons que la loi de Rademacher³ de paramètre $p \in [0, 1]$ est la loi sur \mathbb{R} donnant les poids $1-p$ et p aux valeurs -1 et 1 . Sa moyenne vaut $2p-1$ et sa variance $4p(1-p)$. La loi de Gauss ou loi normale sur \mathbb{R} de moyenne m et de variance σ^2 a pour densité

$$x \mapsto \frac{\exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}.$$

- Figure 1. Les valeurs propres de $\frac{1}{\sqrt{n}}M$ sont approximativement distribuées uniformément sur le disque unité du plan complexe, et ce phénomène est d'autant plus apparent que n est grand. La normalisation en $\frac{1}{\sqrt{n}}$ assure une stabilisation des lignes et des colonnes. En effet, grâce à la loi des grands nombres, la norme de chaque ligne (et de chaque colonne) de $\frac{1}{\sqrt{n}}M$ converge vers $\sqrt{\sigma^2 + |m|^2} = 1$ quand $n \rightarrow \infty$, tandis que leur produit scalaire tend vers $|m|^2 = 0$. Cette structure unitaire asymptotique n'explique pas pourquoi le spectre semble se répartir uniformément sur le disque unité (elle suggérerait plutôt le cercle unité !);
- Figure 2. Les valeurs propres de l'hermitisation multiplicative $\frac{1}{\sqrt{n}}\sqrt{MM^*}$ de $\frac{1}{\sqrt{n}}M$ se distribuent approximativement comme la loi du quart de cercle sur l'intervalle $[0, 2]$, et cela est d'autant plus apparent que la dimension n est grande;
- Figure 3. Les valeurs propres de l'hermitisation additive⁴ $\frac{1}{\sqrt{n}}\frac{M+M^*}{\sqrt{2}}$ de $\frac{1}{\sqrt{n}}M$ se distribuent approximativement comme la loi du demi-cercle sur l'intervalle $[-2, 2]$, et cela est d'autant plus clair que la dimension n est grande, cf. figure 4.

Ces expériences numériques permettent de dégager trois phénomènes d'universalité en grande dimension, capturés mathématiquement par les théorèmes 1.1-1.2-1.3 ci-dessous. Soulignons au passage que les hypothèses sont faites ici sur les coefficients de la matrice, tandis que les résultats concernent les valeurs propres, qui sont des fonctions non-linéaires et non-explicites des coefficients de la matrice.

2. Signifie «indépendants et identiquement distribués» ou «independent and identically distributed».

3. Le terme «Bernoulli» est plutôt réservé à la loi du même type sur $\{0, 1\}$ plutôt que sur $\{-1, 1\}$.

4. La normalisation par $\sqrt{2}$ rend les coefficients hors diagonale de variance 1.

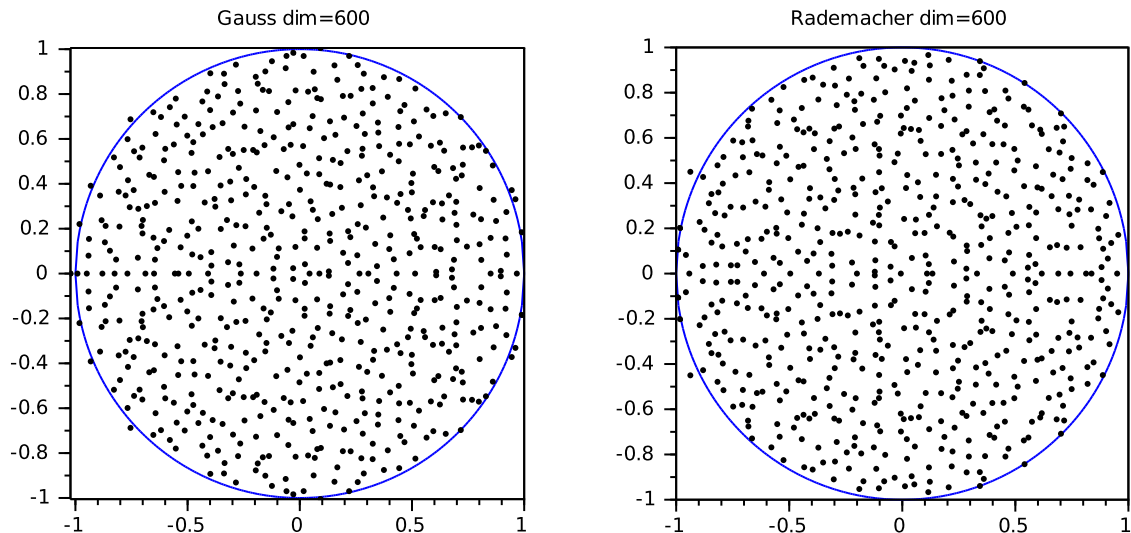


Figure 1 – Phénomène d’universalité de Girko. En noir les valeurs propres de $\frac{1}{\sqrt{n}}M$ où M est une matrice aléatoire carrée $n \times n$ à coefficients i.i.d. de moyenne 0 et de variance 1 (de Gauss à gauche et de Rademacher à droite). En bleu le cercle unité du plan complexe. Note : il est également possible d’utiliser un histogramme en 3D.

```

function girko(dim) // Fonction Scilab.
G = grand(dim,dim,'nor',0,1); // matrice de coef. de Gauss standards
B = sign(G); // matrice de coef. de Rademacher +/-1
sG = spec(G/sqrt(dim)); // spectre et normalisation en racine de dim
sB = spec(B/sqrt(dim)); // idem pour le cas Rademacher
clf(); I = [-1:0.01:1]; J = sqrt(1-I.^2);
subplot(1,2,1); title(sprintf('Gauss_dim=%i',dim));
set(gca(),"tight_limits","on"); set(gca(),"isoview","on");
plot(I,J,2); plot(I,-J,2);
plot(real(sG),imag(sG),'Marker','.', 'MarkerSize',3,'LineStyle','none');
subplot(1,2,2); title(sprintf('Rademacher_dim=%i', dim));
set(gca(),"tight_limits","on"); set(gca(),"isoview","on");
plot(I,J,2); plot(I,-J,2);
plot(real(sB),imag(sB),'Marker','.', 'MarkerSize',3,'LineStyle','none');
endfunction

```

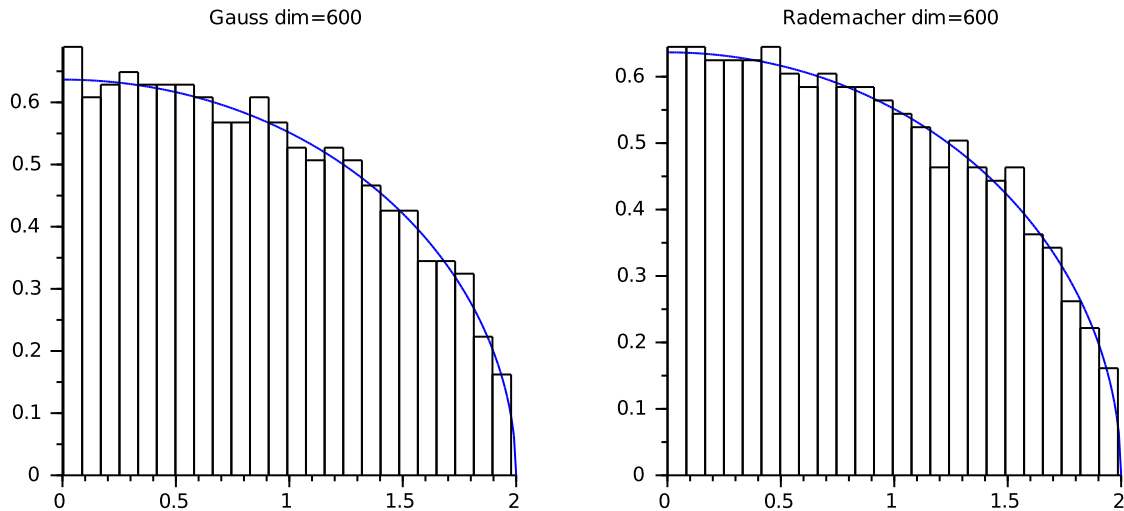


Figure 2 – Phénomène d’universalité de Marchenko-Pastur. En noir, l’histogramme des valeurs propres de $\frac{1}{\sqrt{n}}\sqrt{MM^*}$ où M est une matrice aléatoire carrée $n \times n$ à coefficients i.i.d. de moyenne 0 et de variance 1 (de Gauss à gauche et de Rademacher à droite). En bleu, la densité de la loi du quart de cercle sur l’intervalle $[0, 2]$. Note : la matrice $\frac{1}{\sqrt{n}}\sqrt{MM^*}$ est hermitienne semi-définie positive : son spectre est dans \mathbb{R}_+ .

```

function marpas(dim) // Fonction Scilab.
G = grand(dim,dim,'nor',0,1); // matrice de coef. de Gauss standards
B = sign(G); // matrices de coef. de Rademacher +/-1
sG = sqrt(spec(G*G')/dim); // symétrisation, normalisation, spectre
sB = sqrt(spec(B*B')/dim); // idem pour la version Rademacher
clf(); I = [0:0.01:2]; J = sqrt(4-I.^2)/%pi;
subplot(1,2,1); title(sprintf('Gauss_dim=%i',dim));
set(gca(),"tight_limits","on");
plot2d(I,J,2); histplot(fix(sqrt(dim)),sG);
subplot(1,2,2); title(sprintf('Rademacher_dim=%i', dim));
set(gca(),"tight_limits","on");
plot2d(I,J,2); histplot(fix(sqrt(dim)),sB);
endfunction

```

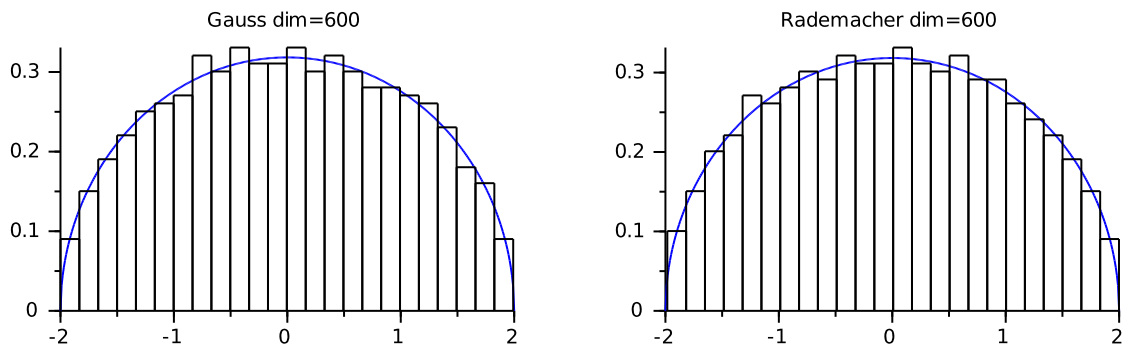


Figure 3 – Phénomène d’universalité de Wigner. En noir l’histogramme des valeurs propres de $\frac{1}{\sqrt{n}} \frac{M+M^*}{\sqrt{2}}$ où M est une matrice aléatoire carrée $n \times n$ à coefficients i.i.d. de moyenne 0 et de variance 1 (de Gauss à gauche et de Rademacher à droite). En bleu la densité de la loi du demi-cercle sur l’intervalle $[-2, 2]$. Note : la matrice $\frac{1}{\sqrt{n}} \frac{M+M^*}{\sqrt{2}}$ est hermitienne, et son spectre est donc réel.

```

function wigner(dim) // Fonction Scilab.
  G = grand(dim,dim,'nor',0,1); // matrice de coef. de Gauss standards
  B = sign(G); // matrices de coef. de Rademacher +/-1
  sG = spec((G+G')/sqrt(2))/sqrt(dim); // symétrisation, normalisation, spectre
  sB = spec((B+B')/sqrt(2))/sqrt(dim); // idem pour la version Rademacher
  clf(); I = [-2:0.01:2]; J = sqrt(4-I.^2)/(2*pi);
  subplot(1,2,1); title(sprintf('Gauss_dim=%i',dim));
  set(gca(),"tight_limits","on");
  plot2d(I,J,2); histplot(fix(sqrt(dim)),sG);
  subplot(1,2,2); title(sprintf('Rademacher_dim=%i', dim));
  set(gca(),"tight_limits","on");
  plot2d(I,J,2); histplot(fix(sqrt(dim)),sB);
endfunction

```

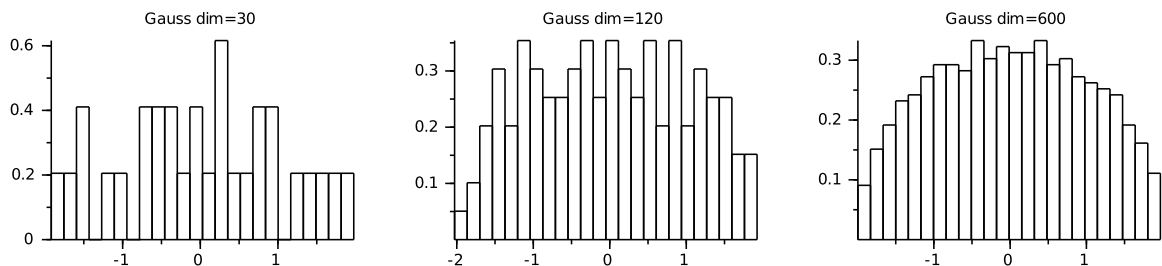


Figure 4 – Graphiques illustrant l’effet de la dimension sur la distribution du spectre des matrices de Wigner (figure 3 et théorème 1.3), pour les dimensions 30, 120, 600. Ces graphiques correspondent à une seule matrice à chaque fois. Si on répète l’expérience, le graphique en petite dimension (à gauche) va beaucoup fluctuer, tandis que celui en grande dimension (à droite) va très peu fluctuer. Un caractère aléatoire s’efface en grande dimension et laisse place à un déterminisme : la loi du demi-cercle.

Afin de décrire mathématiquement la répartition des valeurs propres, on introduit la distribution spectrale empirique d'une matrice $A \in \mathcal{M}_n(\mathbb{C})$,

$$\mu_A := \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k(A)}.$$

Il s'agit tout simplement d'une loi de probabilité discrète, plus précisément d'une mesure de comptage normalisée. Pour tout ensemble $E \subset \mathbb{C}$, la quantité

$$\mu_A(E) = \frac{\text{card}\{1 \leq k \leq n : \lambda_k(A) \in E\}}{n}$$

est la proportion de valeurs propres de A qui sont dans E . Pour toute fonction f on a

$$\int f d\mu_A = \mathbb{E}(f(X)) = \frac{1}{n} \sum_{k=1}^n f(\lambda_k(A))$$

où X est une variable aléatoire discrète de loi μ_A . On retrouve $\mu(E)$ lorsque $f = \mathbf{1}_E$. Si A est aléatoire alors μ_A est une loi de probabilité discrète aléatoire (il y a deux aléas!), et dans ce cas, $\mu_A(E)$ et $\int f d\mu_A$ sont des variables aléatoires. Rappelons qu'on dit qu'une propriété aléatoire a lieu *presque sûrement* lorsqu'elle a lieu sur un événement de probabilité 1. Rappelons que pour tout $n \geq 1$ fixé, M est une matrice aléatoire dont les coefficients $(M_{ij})_{1 \leq i, j \leq n}$ sont i.i.d. de moyenne m et de variance σ^2 .

Théorème 1.1 (Universalité de Girko et loi du cercle). *Presque sûrement, pour tout pavé ou boule E de \mathbb{R}^2 , on a*

$$\lim_{n \rightarrow \infty} \mu_{\frac{1}{\sqrt{n}}M}(E) = \mu(E)$$

où μ est la loi uniforme sur le disque de \mathbb{R}^2 de rayon σ («loi du cercle») de densité

$$(x, y) \mapsto \frac{1}{\pi\sigma^2} \mathbf{1}_{\{(x,y) \in \mathbb{R}^2 : \sqrt{x^2+y^2} \leq \sigma\}}(x, y).$$

Théorème 1.2 (Universalité de Marchenko-Pastur et loi du quart de cercle). *En notant $H := \sqrt{MM^*}$, on a que presque sûrement, pour tout intervalle I de \mathbb{R}_+ ,*

$$\lim_{n \rightarrow \infty} \mu_{\frac{1}{\sqrt{n}}H}(I) = \mu(I)$$

où μ est la loi du quart de cercle sur $[0, 2\sigma]$ et de densité

$$x \mapsto \frac{\sqrt{4\sigma^2 - x^2}}{\pi\sigma^2} \mathbf{1}_{[0, 2\sigma]}(x).$$

Théorème 1.3 (Universalité de Wigner et loi du demi-cercle). *En notant $H := \frac{M+M^*}{\sqrt{2}}$, on a que presque sûrement, pour tout intervalle I de \mathbb{R} ,*

$$\lim_{n \rightarrow \infty} \mu_{\frac{1}{\sqrt{n}}H}(I) = \mu(I)$$

où μ est la loi du demi cercle sur $[-2\sigma, 2\sigma]$ et de densité

$$x \mapsto \frac{\sqrt{4\sigma^2 - x^2}}{2\pi\sigma^2} \mathbf{1}_{[-2\sigma, 2\sigma]}(x).$$

Dans chacun de ces trois théorèmes, presque sûrement, une moyenne empirique aléatoire converge vers une limite déterministe. Cela fait penser à la loi forte des grands nombres⁵, qui ne s'applique pas ici car les valeurs propres ne sont pas indépendantes. D'un autre côté, la limite est décrite par une loi de probabilité universelle qui ne dépend de la loi des coefficients de la matrice de départ qu'à travers leur variance. Cela fait penser au théorème central limite⁶ et sa loi normale universelle.

Si une variable aléatoire de \mathbb{C} (vecteur aléatoire de \mathbb{R}^2) suit la loi uniforme sur le disque unité, de densité $z \mapsto \pi^{-1} \mathbf{1}_{\{z \in \mathbb{C}; |z| \leq 1\}}$, alors la partie réelle et la partie imaginaire suivent toutes les deux la loi du demi-cercle de densité $t \mapsto 2\pi^{-1} \sqrt{1-t^2} \mathbf{1}_{[-1,1]}(t)$. Elles ne sont pas indépendantes. Le rayon suit la loi de densité $r \mapsto r \mathbf{1}_{[0,1]}(r)$ (ce n'est ni la loi uniforme ni la loi du quart de cercle) tandis que la phase suit la loi uniforme sur $[0, 2\pi]$. La valeur absolue de la partie réelle et la valeur absolue de la partie imaginaire suivent la loi du quart de cercle sur $[0, 1]$, de densité $t \mapsto 4\pi^{-1} \sqrt{1-t^2} \mathbf{1}_{[0,1]}(t)$. Malgré ces relations, il n'y a pas de liaison directe entre le spectre de M du théorème 1.1 et le spectre de H des théorèmes 1.2 et 1.3. On observe même un facteur 2 entre le rayon du disque du théorème 1.1 et le rayon des intervalles des théorèmes 1.2 et 1.3. Cela suggère qu'asymptotiquement, la norme d'opérateur $\max_{1 \leq k \leq n} \lambda_k(\sqrt{MM^*})$ de M est deux fois plus grande que le rayon spectral $\max_{1 \leq k \leq n} |\lambda_k(M)|$ de M , ce qui peut être prouvé rigoureusement lorsque $\mathbb{E}(M_{ij}) = 0$ et $\mathbb{E}(|M_{ij}|^4) < \infty$. C'est le signe de l'anormalité de M , cf. section 3.4.

2 Esquisses de preuves du théorème de Wigner

Le théorème 1.3 a été dégagé par Eugene Wigner dans les années 1950. La version considérée ici est due à Leonid Pastur et date des années 1970. Nous esquissons ci-dessous une preuve du théorème 1.3 par la méthode des moments, puis par la méthode de la résolvante. Nous esquissons également deux preuves pour le cas gaussien du GUE, basées sur une minimisation d'énergie, et sur des polynômes orthogonaux.

La matrice aléatoire H est hermitienne, et les variables aléatoires $\{H_{ij}\}_{1 \leq i \leq j \leq n}$ sont indépendantes de moyenne $\sqrt{2}\Re m$. De plus $\{H_{ij}\}_{1 \leq i < j \leq n}$ sont de même loi sur \mathbb{C} de variance σ^2 , tandis que $\{H_{ii}\}_{1 \leq i \leq n}$ sont de même loi sur \mathbb{R} de variance $2\text{Var}(\Re M_{11})$. Nous pouvons dès à présent préciser le pourquoi de la normalisation en $\frac{1}{\sqrt{n}}$: le moment d'ordre 2 de $\mu_{\frac{1}{\sqrt{n}}H}$ se stabilise grâce à la loi forte des grands nombres :

$$\int x^2 d\mu_{\frac{1}{\sqrt{n}}H}(x) = \frac{1}{n} \sum_{k=1}^n \frac{\lambda_k(H)^2}{n} = \frac{1}{n^2} \text{Tr}(H^2) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |H_{ij}|^2 \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \sigma^2 + |m|^2.$$

Le comportement du premier moment est aussi facile à étudier :

$$\int x d\mu_{\frac{1}{\sqrt{n}}H}(x) = \frac{1}{n} \sum_{k=1}^n \frac{\lambda_k(H)}{\sqrt{n}} = \frac{1}{n^{3/2}} \text{Tr}(H) = \frac{H_{11} + \dots + H_{nn}}{n^{3/2}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

Les deux premiers moments de la loi du demi-cercle μ sur $[-2\sigma, 2\sigma]$ valent 0 et σ^2 . Ainsi, le premier moment de $\mu_{\frac{1}{\sqrt{n}}H}$ converge vers le premier moment de μ , tandis que

5. Si $(X_n)_{n \geq 1}$ sont des v.a.r. i.i.d. de moyenne m alors p.s. $\bar{X}_n := \frac{X_1 + \dots + X_n}{n} \rightarrow m$ quand $n \rightarrow \infty$.

6. Si $(X_n)_{n \geq 1}$ sont des v.a.r. i.i.d. de moyenne m et de variance σ^2 alors $\mathbb{P}(\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \in I) \rightarrow \mu(I)$ quand $n \rightarrow \infty$, pour tout intervalle $I \subset \mathbb{R}$, où μ est la loi gaussienne de moyenne 0 et de variance 1.

le second moment de $\mu_{\frac{1}{\sqrt{n}}H}$ ne converge vers celui de μ que si $m = 0$. Cela est dû au comportement explosif de la plus grande valeur propre de $\frac{1}{\sqrt{n}}H$ lorsque $m \neq 0$. Il n'y a pas de contradiction : la convergence en loi de $\mu_{\frac{1}{\sqrt{n}}H}$ n'empêche pas une fraction asymptotiquement négligeable de valeurs propres d'exploser quand $n \rightarrow \infty$.

Modèle réduit

Pour établir le théorème 1.3, une première étape préparatoire, omise ici⁷, consiste à réduire le problème à l'énoncé suivant : soit H une matrice hermitienne $n \times n$ aléatoire telle que $\{H_{ij}\}_{1 \leq i \leq j \leq n}$ sont indépendantes de moyenne 0 et de variance 1, bornées et à valeurs dans $\{z \in \mathbb{C} : |z| \leq C\}$, $C > 0$, avec $\{H_{ii}\}_{1 \leq i \leq n}$ i.i.d. sur \mathbb{R} et $\{H_{ij}\}_{1 \leq i < j \leq n}$ i.i.d. sur \mathbb{C} . Soit μ la loi du demi-cercle sur l'intervalle $[-2, 2]$, de densité $x \mapsto \frac{\sqrt{4-x^2}}{2\pi} \mathbf{1}_{[-2,2]}(x)$. Alors $\mathbb{E}\mu_n := \mathbb{E}\mu_{\frac{1}{\sqrt{n}}H}$ converge en loi quand $n \rightarrow \infty$ vers μ , i.e.

$$\lim_{n \rightarrow \infty} \mathbb{E} \int f d\mu_n = \int f d\mu,$$

pour toute fonction f dans une classe de fonctions suffisamment riche pour garantir la convergence en loi. Nous avons à présent deux méthodes classiques pour achever la preuve du théorème 1.3 : la méthode des moments, de nature combinatoire, basée sur la classe de fonctions $\{x \mapsto x^r : r \in \mathbb{N}\}$, et la méthode de la trace de la résolvante, de nature analytique, basée sur la classe de fonctions $\{x \mapsto (x-z)^{-1} : z \in \mathbb{C}, \Im z > 0\}$.

2.1 Méthode des moments (combinatoire)

On se propose d'établir la convergence des moments pour le modèle réduit :

$$\forall r \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} \mathbb{E} \int x^r d\mu_n(x) \rightarrow \int x^r d\mu.$$

Les moments impairs de μ sont nuls tandis que les pairs sont les nombres de Catalan :

$$\forall r \in \mathbb{N}, \quad \int x^r d\mu(x) = \int_0^2 x^r \frac{\sqrt{4-x^2}}{2\pi} dx = \begin{cases} 0 & \text{si } r = 2k+1, \\ \frac{1}{k+1} \binom{2k}{k} & \text{si } r = 2k. \end{cases}$$

Nous pouvons exprimer les moments de $\mathbb{E}\mu_n$ en fonction des moments joints des coefficients de H grâce à la formule suivante :

$$\mathbb{E} \int x^r d\mu_n(x) = \frac{1}{n^{1+r/2}} \mathbb{E} \sum_{k=1}^n \lambda_k(H)^r = \frac{1}{n^{1+r/2}} \mathbb{E} \text{Tr}(H^r) = \frac{1}{n^{1+r/2}} \sum_{1 \leq i_1, \dots, i_r \leq n} \mathbb{E}(H_{i_1, i_2} \cdots H_{i_r, i_{r+1}})$$

où $i_{r+1} := i_1$. Pour $r = 1$ on obtient bien

$$\mathbb{E} \int x d\mu_n(x) = \frac{1}{n} \mathbb{E} \sum_{i=1}^n H_{ii} = m = 0$$

7. Cette étape fait appel à divers outils : formules variationnelles min-max de Courant-Fischer, inégalité de Hoffman-Wielandt, inégalité de concentration de Azuma-Hoeffding, lemme de Borel-Cantelli, loi forte des grands nombres, fonctions caractéristiques, etc. Certains sont abordés dans l'appendice A.

et pour $r = 2$,

$$\mathbb{E} \int x^2 d\mu_n(x) = \frac{1}{n^2} \mathbb{E} \sum_{i=1}^n \sum_{j=1}^n |H_{ij}|^2 = \sigma^2 + m^2 = 1 = \frac{1}{1+1} \binom{2}{1}.$$

L'étude du cas $r = 3$ est un peu plus subtile. On a

$$\mathbb{E} \int x^3 d\mu_n(x) = \frac{1}{n^{1+3/2}} \sum_{1 \leq i, j, k \leq n} \mathbb{E}(H_{ij}H_{jk}H_{ki}).$$

Si deux éléments parmi $\{\{i, j\}, \{j, k\}, \{k, i\}\}$ sont distincts alors $\mathbb{E}(H_{ij}H_{jk}H_{ki}) = 0$ par indépendance et centrage. Dans le cas contraire, on a $i = k$ ou $i = j$ ou $k = j$. Si $i = k$ alors $\mathbb{E}(H_{ij}H_{jk}H_{ki}) = \mathbb{E}(|H_{ii}|^2 H_{ij})$ qui est nul si $j \neq i$ et qui n'a pas de contribution asymptotique si $j = i$ car on a $n = o(n^{1+3/2})$ triplets vérifiant $i = j = k$. Si $i = j$ alors $\mathbb{E}(H_{ij}H_{jk}H_{ki}) = \mathbb{E}(H_{ii}|H_{ik}|^2)$ et le raisonnement est identique. Si $k = j$ alors $\mathbb{E}(H_{ij}H_{jk}H_{ki}) = \mathbb{E}(|H_{ij}|^2 H_{jj})$ et le raisonnement est à nouveau identique. Ainsi, le moment d'ordre 3 de $\mathbb{E}\mu_n$ tends vers 0 quand $n \rightarrow \infty$.

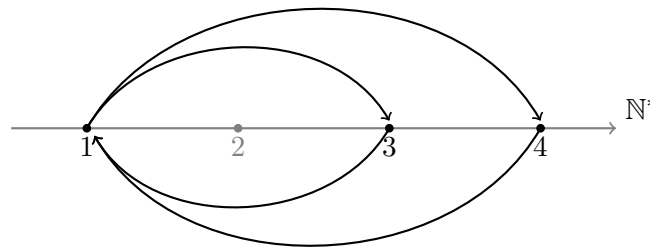


Figure 5 – Graphe pour $r = 4$, $i_1 = i_3 = 1$, $i_2 = 4$, $i_4 = 3$, correspondant à $\mathbb{E}(H_{14}H_{41}H_{13}H_{31})$. Les arêtes sont $1 \rightarrow 4$, $4 \rightarrow 1$, $1 \rightarrow 3$, $3 \rightarrow 1$. On a $t = 3$.

Lorsque r est quelconque et **fixé**, on associe à chaque r -uplet d'indices i_1, \dots, i_r un graphe orienté, obtenu en positionnant leurs valeurs sur une demi-droite horizontale figurant \mathbb{N}^* , cf. figure 5. Les sommets du graphe sont les valeurs distinctes prises par ces indices, et on note t leur nombre. Les arêtes du graphe sont les liaisons (i_k, i_{k+1}) avec $1 \leq k \leq r+1$ et $i_{r+1} := i_1$. On dit qu'il s'agit d'un graphe $G(t)$. Deux graphes $G(t)$ sont équivalents lorsque qu'on peut passer de l'un à l'autre en permutant les indices. Des graphes $G(t)$ équivalents donnent la même valeur à $\mathbb{E}(H_{i_1, i_2} \cdots H_{i_r, i_{r+1}})$, notée $\mathbb{E}(H_G)$. Il y a $n(n-1) \cdots (n-t+1)$ graphes $G(t)$ dans chaque classe d'équivalence. Par ailleurs, dans chaque classe d'équivalence, on peut supposer que les sommets sont $1, \dots, t \leq r$, et ainsi le nombre total de classes d'équivalence c_r ne dépend que de r . Afin de calculer les contributions, on distingue trois types de classes de graphes $G(t)$:

- *type 1* : ceux pour qui chaque arête est présente dans l'autre sens le même nombre de fois, et le graphe squelette obtenu en effaçant les orientations des arêtes est un graphe sans cycles, c'est-à-dire un arbre. Exemple : le cas de la figure 5, qui donne l'arbre à trois sommets dont les arêtes sont $1 \leftrightarrow 3$ et $1 \leftrightarrow 4$;
- *type 2* : ceux pour qui une arête au moins n'apparaît qu'une seule fois. Exemple : $\mathbb{E}(H_{12}H_{23}H_{31})$ ne contient qu'une seule fois les arêtes $1 \leftrightarrow 2$, $2 \leftrightarrow 3$, $3 \leftrightarrow 1$;

- *type 3* : ceux qui ne sont pas de type 1 ou de type 2. C’est le cas par exemple pour $\mathbb{E}(H_{12}H_{23}H_{31}H_{12}H_{23}H_{31})$, qui n’est ni de type 1 ni de type 2 car les arêtes $1 \leftrightarrow 2, 2 \leftrightarrow 3, 3 \leftrightarrow 1$ apparaissent chacune deux fois dans le même sens, ou encore pour $\mathbb{E}(H_{12}H_{21}H_{13}H_{32}H_{23}H_{31})$, qui n’est ni de type 1 ni de type 2 car les arêtes $1 \leftrightarrow 2, 2 \leftrightarrow 3, 3 \leftrightarrow 1$ apparaissent chacune dans les deux sens mais le graphe obtenu en effaçant les orientations des arêtes est le cycle $1 \leftrightarrow 2 \leftrightarrow 3 \leftrightarrow 1$.

Nous disposons des faits suivants :

- *Les graphes de type 2 ont une contribution nulle.* En effet, si G est un graphe de type 2 alors $\mathbb{E}(H_G) = 0$ par indépendance et centrage ;
- *Les graphes de type 3 ont une contribution asymptotiquement nulle.* En effet, si G est un graphe $G(t)$ de type 3 alors on peut établir sans grande difficulté que $t \leq (r+1)/2$, qui donne $n(n-1)\cdots(n-t+1) \leq n^t \leq n^{1/2+r/2}$. D’autre part, le nombre de classes de type 3 est majoré par le nombre total de classes $c_r = \mathcal{O}(1)$. À présent, comme les coefficients de H sont bornés, on a $\mathbb{E}(H_G) = \mathcal{O}(1)$, et donc

$$\sum_{\text{classes t.3}} \frac{n(n-1)\cdots(n-t+1)}{n^{1+r/2}} \mathbb{E}(H_G) = \mathcal{O}(n^{-1/2}) = o_{n \rightarrow \infty}(1);$$

- *Seuls les graphes de type 1 contribuent asymptotiquement.* En effet, pour tout graphe G de type 1 on a $\mathbb{E}(H_G) = 1$ par indépendance car les coefficients de H ont une variance de 1. Cela ramène le problème à la détermination du nombre de classes de type 1. Il n’y a pas de graphes de type 1 si r est impair. Si r est pair, disons $r = 2k$, alors $t = r/2$, et les classes de type 1 sont en bijection avec les parenthésages : il y en a donc $\frac{1}{k+1} \binom{2k}{k}$ (nombres de Catalan), d’où enfin

$$\sum_{\text{cl. t.1}} \frac{n(n-1)\cdots(n-t+1)}{n^{1+r/2}} \mathbb{E}(H_G) = \frac{n}{n} \cdots \frac{n-k+1}{n} \frac{1}{k+1} \binom{2k}{k} \xrightarrow{n \rightarrow \infty} \frac{1}{1+k} \binom{2k}{k}.$$

Ceci achève notre esquisse de preuve du théorème 1.3 par la méthode des moments. En plus des parenthésages, les nombres de Catalan permettent de dénombrer des suites d’objets variées comme les excursions de la marche aléatoire simple, les partitions non-croisées, les chemins de Dyck, etc, qui sont donc toutes en bijection. Signalons par ailleurs que lorsque M est gaussienne, il est possible d’exprimer les moments avec la matrice de covariance en utilisant une formule de Wick, qui est une conséquence de l’intégration par parties, cf. [G1].

2.2 Méthode de la résolvante (analyse)

Nous venons de voir que la famille de fonctions test $\{x \mapsto x^r : r \in \mathbb{N}\}$ ramène le problème à la combinatoire des moments. Soit $\mathbb{C}_+ := \{z \in \mathbb{C} : \Im z > 0\}$. Il se trouve que la famille de fonctions test $\{x \mapsto 1/(x-z) : z \in \mathbb{C}_+\}$ ramène le problème à une équation pour la trace de la résolvante, liée à la transformée de Cauchy-Stieltjes.

La transformée de Cauchy-Stieltjes $S_\nu : \mathbb{C}_+ \rightarrow \mathbb{C}$ d’une loi ν sur \mathbb{R} est définie par

$$S_\nu(z) := \int \frac{1}{x-z} d\nu(x).$$

On a toujours $|S_\nu(z)| \leq 1/\Im z$ et $\Im S_\nu(z) \geq 0$. La fonction S_ν caractérise ν . Elle est analytique sur \mathbb{C}_+ . De plus, une suite $(\nu_n)_{n \geq 1}$ de lois sur \mathbb{R} converge en loi vers une loi ν sur \mathbb{R} si et seulement si $(S_{\nu_n})_{n \geq 1}$ converge ponctuellement vers S_ν sur \mathbb{C}_+ . Si

$\nu_n = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k(A)}$ est la distribution spectrale empirique d'une matrice hermitienne A , et si $G_A(z) := (A - zI)^{-1}$ est la résolvante de A en $z \in \mathbb{C}_+$, alors

$$S_{\nu_n}(z) = \frac{1}{n} \sum_{k=1}^n \frac{1}{\lambda_k(A) - z} = \frac{1}{n} \sum_{k=1}^n \frac{1}{\lambda_k(A - zI)} = \frac{1}{n} \text{Tr}(G_A(z)).$$

Si A est aléatoire, on a $\mathbb{E}(S_{\nu_n}(z)) = S_{\mathbb{E}\nu_n}(z)$ grâce au théorème de Fubini-Tonelli.

Pour établir le théorème 1.3, on observe tout d'abord que si μ est la loi du demi-cercle de densité $x \mapsto \frac{\sqrt{4-x^2}}{2\pi} \mathbf{1}_{[-2,2]}(x)$ alors un calcul par la méthode des résidus donne

$$S_\mu(z) = \frac{-z + \sqrt{z^2 - 4}}{2},$$

en fait l'unique solution à partie imaginaire positive de l'équation du second degré $s^2 + sz + 1 = 0$ en s , qui s'écrit également, sous forme de point fixe,

$$s + \frac{1}{z + s} = 0.$$

Soit μ_n la mesure spectrale empirique de $\frac{1}{\sqrt{n}}H_n$. L'idée est d'établir que $S_{\mathbb{E}\mu_n}$ est solution approchée de cette équation. La formule d'inversion par blocs de Schur affirme que pour toute partition $\{1, \dots, n\} = K \cup K^c$,

$$(A^{-1})_{K,K} = (A_{K,K} - A_{K,K^c}(A_{K^c,K^c})^{-1}A_{K^c,K})^{-1}.$$

En utilisant cette formule pour la résolvante de $\frac{1}{\sqrt{n}}H$, avec $K = \{k\}$ où k varie de 1 à n , on obtient, après de longs calculs,

$$\varepsilon_n := S_{\mathbb{E}\mu_n}(z) + \frac{1}{z + S_{\mathbb{E}\mu_n}(z)} \rightarrow 0.$$

Cela conduit au résultat. Les détails, omis ici, et qu'on peut trouver par exemple dans [BS] ou [AGZ], ne sont pas plus simples que pour la méthode des moments. Signalons par ailleurs que lorsque M est gaussienne, il est possible d'utiliser l'identité de la résolvante $G_B - G_A = G_A(A - B)G_B$ et une intégration par parties plutôt que la formule d'inversion par blocs, et cette stratégie s'étend même au delà du cas gaussien par une technique d'interpolation, cf. [PS].

2.3 Méthode de minimisation d'énergie pour GUE

On considère dans cette section le cas gaussien où $\{\Re M_{ij}, \Im M_{ij}\}_{1 \leq i, j \leq n}$ sont i.i.d. de loi gaussienne de moyenne 0 et de variance 1/2. Dans ce cas, la loi de la matrice hermitienne aléatoire $H = (M + M^*)/\sqrt{2}$ du théorème 1.3 vérifie :

- $\{H_{ii}, \Re H_{ij}, \Im H_{ij}\}_{1 \leq i < j \leq n}$ indépendantes et de loi gaussienne centrée ;
- $\{H_{ii}\}_{1 \leq i \leq n}$ de variance 1 ;
- $\{\Re H_{ij}, \Im H_{ij}\}_{1 \leq i < j \leq n}$ de variance 1/2, et donc $\{H_{ij}\}_{1 \leq i < j \leq n}$ de variance 1.

Identifions l'ensemble des matrices hermitiennes $\mathcal{H}_n := \{H \in \mathcal{M}_n(\mathbb{C}) : H^* = H\}$ à $\mathbb{C}^{(n^2-n)/2} \times \mathbb{R}^n \equiv \mathbb{R}^{n^2}$. La loi de H admet une densité proportionnelle à

$$H \mapsto e^{-\frac{1}{2} \sum_{i,j=1}^n |H_{ij}|^2} = e^{-\frac{1}{2} \text{Tr}(HH^*)} = e^{-\frac{1}{2} \text{Tr}(H^2)}.$$

La loi de H est invariante par conjugaison unitaire : UHU^* et H ont même loi pour toute matrice unitaire U . On dit que H appartient à l'ensemble gaussien unitaire,

en anglais « Gaussian Unitary Ensemble » (GUE). Le changement de variables $H \leftrightarrow (U, D)$ issu de la diagonalisation $H = U^*DU$ en base orthonormée permet d'établir, par intégration de la variable U , que les valeurs propres $\lambda_1(\frac{1}{\sqrt{n}}H), \dots, \lambda_n(\frac{1}{\sqrt{n}}H)$ de $\frac{1}{\sqrt{n}}H$ admettent une densité proportionnelle à

$$\lambda \in \mathbb{R}^n \mapsto e^{-\frac{n}{2} \sum_{i=1}^n \lambda_i^2} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^2. \quad (*)$$

Le déterminant de Vandermonde ci-dessus exprime une dépendance, et provient du jacobien du changement de variables. La densité s'écrit également $e^{-E_n(\lambda)}$ où

$$E_n(\lambda) := \frac{n}{2} \sum_{i=1}^n \lambda_i^2 + 2 \sum_{1 \leq i < j \leq n} \ln \frac{1}{|\lambda_i - \lambda_j|}.$$

Cela permet de concevoir $\lambda_1(\frac{1}{\sqrt{n}}H), \dots, \lambda_n(\frac{1}{\sqrt{n}}H)$ comme des particules chargées dans le plan, confinées sur la droite réelle, dont l'énergie de configuration est $E_n(\lambda)$. La structure de E_n indique que ces particules subissent à la fois un champ extérieur quadratique qui les localise près de l'origine, ainsi qu'une force coulombienne qui les fait se repousser mutuellement. De ces contraintes antagonistes naît un équilibre macroscopique quand $n \rightarrow \infty$. Réécrivons l'énergie $E_n(\lambda)$ au moyen de la mesure de comptage des valeurs propres notée $\mu_n := \mu_{\frac{1}{\sqrt{n}}H}$:

$$E_n(\lambda) = n^2 \left(\int \frac{x^2}{2} d\mu_n(x) + \iint_{x \neq y} \ln \frac{1}{|x - y|} d\mu_n(x) d\mu_n(y) \right).$$

Ainsi, la loi du spectre de $\frac{1}{\sqrt{n}}H$ est proportionnelle à $e^{-E_n(\lambda_1, \dots, \lambda_n)} \approx e^{-n^2 E(\mu_n)}$ où E est la fonctionnelle quadratique strictement convexe définie pour toute loi ν sur \mathbb{R} par

$$E(\nu) := \int \frac{x^2}{2} d\nu(x) + \iint \ln \frac{1}{|x - y|} d\nu(x) d\nu(y).$$

Cela suggère que quand $n \rightarrow \infty$, la mesure empirique μ_n converge vers l'ensemble des minima de la fonctionnelle E . Or on peut établir précisément que $\nu \mapsto E(\nu)$ est minimisée par une unique loi de probabilité : la loi du demi-cercle μ de densité

$$x \mapsto \frac{\sqrt{4 - x^2}}{2\pi} \mathbf{1}_{[-2, 2]}(x).$$

Cela permet d'établir le théorème 1.3 dans le cas gaussien du GUE. Cette méthode, détaillée dans [AGZ], est un exemple spécial de principe de grandes déviations. Elle reste valable au delà du GUE, pour des modèles de matrices aléatoires unitairement invariants de densité proportionnelle à $e^{-n \text{Tr}(V(H))}$, où par définition $\text{Tr}(V(H)) := \sum_{k=1}^n V(\lambda_k(H))$. La loi obtenue à la limite dépend de V et n'est pas la loi du demi-cercle en général, ce qui montre qu'il n'y a pas universalité.

2.4 Méthode des polynômes orthogonaux pour GUE

Soit $(P_k)_{k \geq 0}$ la suite des polynômes de Hermite, orthogonaux et normalisés pour la loi gaussienne standard sur \mathbb{R} de densité $\gamma(x) = (2\pi)^{-1/2} e^{-x^2/2}$. Pour tous $k, k' \in \mathbb{N}$,

$$\int P_k(x) P_{k'}(x) \gamma(x) dx = \mathbf{1}_{k=k'}.$$

Les polynômes $(P_k)_{k \geq 0}$ vérifient la récurrence à trois termes

$$xP_k = \sqrt{k+1}P_{k+1} + \sqrt{k}P_{k-1}.$$

Réécrivons la densité (\star) du GUE en utilisant ces polynômes. Le déterminant de Vandermonde $\prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|$ est proportionnel à $\det((P_{i-1}(\lambda_j))_{1 \leq i, j \leq n})$ (cela découle uniquement du fait que P_k est un polynôme de degré k pour tout k). En posant

$$K(x, y) := \sum_{k=0}^n P_k(x)P_k(y),$$

on obtient que $\prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^2 = \left(\prod_{1 \leq i < j \leq n} |K(\lambda_i, \lambda_j)| \right)^2$ est proportionnel à

$$\det((K(\lambda_i, \lambda_j))_{1 \leq i, j \leq n}).$$

Cela fournit une représentation déterminantale de la densité (\star) du GUE, basée sur le noyau K , point de départ d'une analyse fine. Cette approche, due à Gaudin et Mehta, permet d'établir en particulier que si H est GUE alors la densité de $\mathbb{E}\mu_H$ s'écrit

$$x \mapsto \gamma(x) \frac{1}{n} \sum_{k=0}^{n-1} P_k^2(x).$$

La densité de $\mu_n := \mathbb{E}\mu_{\frac{1}{\sqrt{n}}H}$ s'écrit donc

$$x \mapsto \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} P_k^2(\sqrt{n}x) \gamma(\sqrt{n}x).$$

À partir de cette formule, une première méthode, exploitant les propriétés d'orthogonalité de la suite $(P_k)_{k \geq 0}$, permet d'obtenir une formule de récurrence à trois termes pour les moments de μ_n , qui fait écho à la formule de récurrence à trois termes pour les polynômes orthogonaux $(P_k)_{k \geq 0}$. Cela permet d'établir finalement que μ_n converge vers la loi du demi-cercle par une méthode de moments.

Alternativement, par des calculs utilisant l'intégration par parties et l'orthogonalité de $(P_k)_{k \geq 0}$, on peut établir que la transformée de Laplace au point θ de la loi de densité $x \mapsto P_n^2(\sqrt{n}x) \gamma(\sqrt{n}x)$ est solution quand $n \rightarrow \infty$ de l'équation différentielle $4\theta f'' + 2f' - \theta f = 0$, qui est précisément l'équation différentielle satisfaite par la transformée de Laplace d'une loi de l'arcsinus sur $[-2, 2]$. Cette observation permet d'établir que la loi de densité $x \mapsto P_n^2(\sqrt{n}x) \gamma(\sqrt{n}x)$ converge en loi quand $n \rightarrow \infty$ vers la loi de l'arcsinus sur $[-2, 2]$ de densité $x \mapsto \pi^{-1}(4-x^2)^{-1/2} \mathbf{1}_{[-2, 2]}(x)$. En exploitant le fait que la densité de μ_n est une moyenne de Cesàro, on peut en déduire que μ_n converge en loi vers la loi de $\sqrt{U}\xi$ où U et ξ sont des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$ et de loi de l'arcsinus sur $[-2, 2]$ respectivement. Il s'agit là d'une représentation particulière de la loi du demi-cercle sur $[-2, 2]$!

Les détails ainsi que de nombreuses références se trouvent dans [L].

3 Au delà du théorème de Wigner

L'étude des matrices aléatoires ne se résume pas à l'étude du spectre. Le comportement des vecteurs propres est par exemple d'une grande importance pour la com-

préhension des modèles de graphes aléatoires, mais aussi des opérateurs de Schrödinger aléatoires en physique en liaison avec le phénomène de localisation d'Anderson. D'autre part, les propriétés non asymptotiques des modèles de matrices aléatoires jouent un rôle important dans les applications, comme par exemple pour les problèmes de complétion, d'échantillonnage, de classification, etc.

On distingue deux grandes familles de modèles de matrices aléatoires. La première famille est constituée par les modèles dans lesquels les coefficients de la matrice sont i.i.d. comme dans les théorèmes 1.1, 1.2, 1.3. Dans ce cas, la loi du spectre et des vecteurs propres n'est en général pas connue. La seconde famille est constituée par les modèles dont la loi est unitairement invariante. Dans ce cas, les vecteurs propres sont indépendants des valeurs propres, la matrice des vecteurs propres suit la loi uniforme sur le groupe unitaire, et la loi des valeurs propres est connue (elle fait souvent intervenir un déterminant de Vandermonde lié au jacobien du changement de variable comme dans l'exemple du GUE). Le modèle gaussien GUE appartient aux deux familles : les coefficients de la matrice sont i.i.d. et la loi est unitairement invariante. Le caractère explicite de la loi des valeurs propres des modèles unitairement invariants comme le GUE permet d'effectuer des calculs précis, ce qui conduit à dégager des comportements quantitatifs ou asymptotiques, dont certains sont universels, en ce sens qu'il restent valables au delà du modèle considéré. Nous renvoyons à [ER] pour un panorama riche et éclairant, en liaison avec les noyaux de Bessel, de Airy, et le noyau sinus, ainsi qu'avec les décompositions matricielles classiques.

3.1 Comportement au bord et universalité locale

Le théorème 1.3 de Wigner concerne une universalité au niveau global (ensemble du spectre) et au premier ordre. Le comportement global au second ordre consiste par exemple à étudier le comportement asymptotique quand $n \rightarrow \infty$ de

$$\int f d\mu_{\frac{1}{\sqrt{n}}H} - \int f d\mu.$$

On peut établir qu'en normalisant par un coefficient qui dépend de n et de f , on obtient une convergence en loi vers une loi gaussienne. On dit qu'il s'agit d'un théorème central limite pour les statistiques linéaires du spectre. Lorsque f est assez régulière, la normalisation ne dépend pas de n , tandis que lorsque f est une indicatrice d'intervalle, la normalisation peut être très différente de \sqrt{n} . Ces comportements inhabituels sont dûs à la dépendance (répulsion) entre les valeurs propres. Notons par ailleurs que pour $f(x) = x^r$ avec $r \in \mathbb{N}^*$, nous retrouvons les moments, qui convergent à vitesse \sqrt{n} pour le modèle réduit. Nous renvoyons à [Da] pour tous ces aspects.

La convergence globale fournie par le théorème 1.3 de Wigner n'empêche pas une fraction asymptotiquement négligeable du spectre de diverger vers l'infini. Si par exemple $m \neq 0$ alors on peut établir qu'une valeur propre va diverger vers l'infini à vitesse \sqrt{n} . Ceci mène à l'étude du comportement du bord du spectre. Au premier ordre, on montre que si $m = 0$ et si le moment d'ordre 4 de H_{12} est fini alors les valeurs propres extrêmes convergent vers le bord du support de la loi limite :

$$\lambda_{\min} := \min_{1 \leq k \leq n} \lambda_k\left(\frac{1}{\sqrt{n}}H\right) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} -2\sigma \quad \text{et} \quad \lambda_{\max} := \max_{1 \leq k \leq n} \lambda_k\left(\frac{1}{\sqrt{n}}H\right) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 2\sigma.$$

Au second ordre, on montre sous les mêmes hypothèses que les variables aléatoires

$$n^{2/3}(-2\sigma - \lambda_{\min}) \quad \text{et} \quad n^{2/3}(\lambda_{\max} - 2\sigma)$$

convergent en loi quand $n \rightarrow \infty$ vers une loi de Tracy-Widom, liée à la fonction d’Airy et à une équation de Painlevé de type II. Sur le modèle du GUE, cela est lié au fait que le noyau K du GUE converge au bord vers un noyau d’Airy. Loin du bord, c’est plutôt un noyau sinus qui apparaît à la limite. Nous renvoyons à [De, ER].

Il est également possible d’étudier le spectre localement en quantifiant par exemple l’écart entre les valeurs propres individuelles et la position théorique liée aux quantiles de la loi limite. Cela peut également consister à comparer quantitativement la loi des valeurs propres à celle du GUE, et à établir l’universalité de la limite du noyau du GUE (noyau sinus). C’est l’objet de travaux récents présentés dans [G2, E].

Une méthode due à Lindeberg permet d’établir le théorème central limite classique pour des variables aléatoires i.i.d. en remplaçant une à une les variables par des variables gaussiennes de même espérance et de même variance, et en contrôlant l’erreur commise. Il se trouve que cette méthode de « remplacement gaussien » permet également de déduire le théorème 1.3 de Wigner à partir du cas gaussien GUE, cf. [T1]. Cela fait appel à une formule de Hadamard qui exprime la dérivée des valeurs propres par rapport aux coefficients de la matrice. En utilisant cette méthode, il est possible d’établir que le comportement local des valeurs propres individuelles ne dépend que des quatre premiers moments des coefficients de la matrice, cf. [Da].

Rappelons que les écarts entre valeurs propres constituaient l’objectif initial de Wigner. Si $\lambda_1 \leq \dots \leq \lambda_n$ sont les valeurs propres ordonnées du GUE (normalisé pour que les coefficients de la matrice soient de variance $1/n$) alors on peut établir que la distribution moyenne des écarts normalisés $n(\lambda_{k+1} - \lambda_k)$ converge en loi quand $n \rightarrow \infty$ vers une loi spéciale de Gaudin-Mehta liée à l’équation différentielle de Painlevé V. Notons que si les valeurs propres étaient par exemple i.i.d. de loi uniforme sur l’intervalle $[-2, 2]$ alors la loi limite des écarts normalisés serait la loi exponentielle de moyenne 4, ce qui n’est pas le cas ici, à cause du phénomène de répulsion, cf. [T2].

3.2 Probabilités libres de Voiculescu

Le phénomène d’universalité le plus classique en théorie des probabilités est celui mis en lumière par le théorème central limite. Il se trouve que la loi du demi-cercle constitue à bien des égards un analogue, pour les grandes matrices aléatoires, de la loi gaussienne, comme l’explique la *théorie des probabilités libres* de Voiculescu.

Probabilités libres

Considérons une algèbre⁸ \mathcal{A} sur \mathbb{C} possédant une unité 1 (élément neutre de la multiplication), munie d’une involution $a \mapsto a^*$, c’est-à-dire que $(a^*)^* = a$ pour tout $a \in \mathcal{A}$, et d’une forme linéaire $\tau : \mathcal{A} \rightarrow \mathbb{C}$ vérifiant $\tau(1) = 1$, $\tau(ab) = \tau(ba)$ et $\tau(aa^*) \geq 0$ pour tout $a \in \mathcal{A}$, avec égalité si et seulement si $a = 0$. Un exemple simple est fourni par $\mathcal{A} = \mathcal{M}_n(\mathbb{C})$ et $a^* = \bar{a}^\top$, et $\tau = \frac{1}{n} \text{Tr}$. Dans cet exemple, pour tout $a \in \mathcal{A}$, et tout $m \in \mathbb{N}$, la quantité $\tau(a^m)$ est le moment d’ordre m de la distribution spectrale de a . Au delà de cet exemple, nous allons conserver cette interprétation de τ comme une « espérance », et considérer que les éléments de \mathcal{A} sont des « variables aléatoires algébriques ». On dit que $a \in \mathcal{A}$ est centré et réduit lorsque $\tau(a) = 0$ et $\tau(a^2) = 1$.

8. L’addition est commutative, mais pas forcément la multiplication.

Moments et loi

On dit que $a \in \mathcal{A}$ est réel lorsque $a = a^*$. On appelle loi de $a \in \mathcal{A}$ la donnée de

$$\tau(a_1 \cdots a_m), \quad m \in \mathbb{N}^*, \quad a_1, \dots, a_m \in \{a, a^*\}.$$

Si a est réel alors sa loi est caractérisée par ses moments $\{\tau(a^r) : r \in \mathbb{N}\}$. Lorsque $\mathcal{A} = \mathcal{M}_n(\mathbb{C})$ et $a \in \mathcal{A}$ est réel, alors a est une matrice hermitienne et sa distribution spectrale empirique $\mu_a = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k(a)}$, qui est une loi de probabilité discrète finie, est caractérisée par ses moments, qui sont exactement $\{\tau(a^r) : r \in \mathbb{N}\}$:

$$\tau(a^r) = \frac{1}{n} \text{Tr}(a^r) = \frac{1}{n} \sum_{i=1}^n \lambda_i(a)^r = \int x^r d\mu_a(x).$$

De ce point de vue, le concept de loi d'un élément de \mathcal{A} est une extension algébrique du concept de distribution spectrale empirique des matrices hermitiennes.

Liberté

Dans la théorie des probabilités classiques sur $(\Omega, \mathcal{F}, \mathbb{P})$, des sous-tribus $\mathcal{F}_1, \dots, \mathcal{F}_m$ de \mathcal{F} sont indépendantes lorsque $\mathbb{E}(X_1 \cdots X_k) = 0$ pour toutes variables aléatoires bornées X_1, \dots, X_m vérifiant $\mathbb{E}(X_k) = 0$ et X_k est \mathcal{F}_k -mesurable pour tout $1 \leq k \leq m$. Nous allons imiter ce concept dans notre cadre algébrique, en tenant compte de la principale différence avec le cadre classique : le caractère non-commutatif.

On dit que des sous-algèbres (contenant 1 et stables par involution) $\mathcal{A}_1, \dots, \mathcal{A}_m$ de \mathcal{A} sont *libres* lorsque $\tau(a_1 \cdots a_k) = 0$ pour tous $k \in \mathbb{N}^*$, tous $i_1, \dots, i_k \in \{1, \dots, m\}$, et tous $a_1 \in \mathcal{A}_{i_1}, \dots, a_k \in \mathcal{A}_{i_k}$ tels que $\tau(a_1) = \dots = \tau(a_k) = 0$ et $i_j \neq i_{j+1}$ pour tout $1 \leq j < k$. On dit que $a, b \in \mathcal{A}$ sont libres lorsque les sous-algèbres qu'ils engendrent sont libres. La liberté permet de calculer des moments joints, de les exprimer complètement en fonction des moments des arguments. Si par exemple $a, b \in \mathcal{A}$ sont libres alors pour obtenir $\tau(ab)$ on écrit grâce à la liberté,

$$0 = \tau((a - \tau(a))(b - \tau(b))) = \tau(ab) - \tau(a)\tau(b)$$

d'où $\tau(ab) = \tau(a)\tau(b)$ (dans ce cas simple, la non-commutativité ne s'exprime pas).

Théorème central limite libre

Nous pouvons à présent énoncer le théorème central limite libre : si $a_1, \dots, a_n \in \mathcal{A}$ sont réels, centrés et réduits, libres, et de même loi, alors la loi de

$$s_n := \frac{a_1 + \cdots + a_n}{\sqrt{n}}$$

converge vers la loi du demi-cercle centrée et de variance 1, autrement dit

$$\forall r \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} \tau(s_n^r) = \begin{cases} 0 & \text{si } r = 2k + 1, \\ \frac{1}{k+1} \binom{2k}{k} & \text{si } r = 2k. \end{cases}$$

La preuve de ce théorème central limite consiste à calculer la loi de s_n . Notons que s_n est centré et réduit, tout comme dans le théorème central limite classique.

Convolution libre

La liberté permet de définir une notion de convolution. Si $a, b \in \mathcal{A}$ sont libres de loi μ_a et μ_b alors la loi μ_{a+b} de $a+b$ ne dépend que de μ_a et de μ_b . On la note $\mu_a \boxplus \mu_b$ et on dit qu'il s'agit de la *convolution libre* de μ_a et μ_b . La famille des lois gaussiennes sur \mathbb{R} engendrée par translation et dilatation à partir de la loi gaussienne standard centrée et réduite, est stable par convolution classique (additivité des variances). Il en va de même pour la famille des lois du demi-cercle sur \mathbb{R} avec la convolution libre (additivité des variances!).

Liberté asymptotique

Si A et B sont deux matrices hermitiennes $n \times n$, alors le spectre de $A+B$ dépend non seulement du spectre de A et du spectre de B , mais aussi des vecteurs propres des A et de B (sauf si A et B commutent car dans ce cas elles sont diagonalisables dans la même base). Si maintenant A et B sont deux matrices aléatoires hermitiennes indépendantes, il n'y a aucune raison que μ_{A+B} ne dépende que de μ_A et μ_B . Le cas du GUE est cependant remarquable. Soit donc A une matrice $n \times n$ aléatoire hermitienne gaussienne du GUE normalisé de sorte que la distribution spectrale empirique μ_A ait un moment d'ordre 2 égal à 1. Alors les moments de $\mathbb{E}\mu_A$ convergent quand $n \rightarrow \infty$ vers ceux de la loi du demi-cercle centrée de variance 1. Si B est une copie indépendante de A , alors $A+B$ suit la même loi que $\sqrt{2}A$, et donc les moments de $\mathbb{E}\mu_{A+B}$ convergent quand $n \rightarrow \infty$ vers ceux de la loi du demi-cercle centrée de variance 2. On a donc en quelque sorte la relation $\mathbb{E}\mu_{A+B} = \mathbb{E}\mu_A \boxplus \mathbb{E}\mu_B$ quand $n \rightarrow \infty$. Voiculescu a établi que ce phénomène de *liberté asymptotique* reste vrai au delà des couples de matrices indépendantes du GUE, à condition de «découpler» aléatoirement les espaces propres des deux matrices par conjugaison unitaire aléatoire. Soient par exemple A et B des matrices hermitiennes $n \times n$ telles que $\mu_A \rightarrow \mu_a$ et $\mu_B \rightarrow \mu_b$ au sens des moments quand $n \rightarrow \infty$, où μ_a et μ_b sont deux lois à support compact. Soient U, V des matrices unitaires aléatoires indépendantes et de loi uniforme (on dit «Haar unitaires»). Alors $\mathbb{E}\mu_{U A U^* + V B V^*}$ converge au sens des moments vers $\mu_a \boxplus \mu_b$.

La théorie des probabilités libres est très riche. Elle apporte une vision algébrique très éclairante sur l'asymptotique des matrices aléatoires de grande dimension. Nous renvoyons à [AGZ] pour plus d'information à ce sujet.

3.3 Quelques mots sur le théorème de Marchenko-Pastur

Nous commençons par une motivation statistique, celle de Wishart. Soit X_1, \dots, X_n des vecteurs colonne aléatoires i.i.d. de \mathbb{R}^d de moyenne 0 et de matrice de covariance Σ . La matrice Σ est symétrique $d \times d$ et ses valeurs propres sont positives ou nulles. On a $\Sigma_{ij} = \mathbb{E}(X_{ki} X_{kj})$ pour tous $1 \leq i, j \leq d$ et tout $1 \leq k \leq n$ ou encore

$$\Sigma = \mathbb{E}(X_1 X_1^\top) = \dots = \mathbb{E}(X_n X_n^\top).$$

La *matrice de covariance empirique* $\widehat{\Sigma}_n$ est la matrice symétrique $d \times d$ définie par

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{k=1}^n X_k X_k^\top = \frac{1}{n} (X_1 \cdots X_n) (X_1 \cdots X_n)^\top.$$

On a $\mathbb{E}(\widehat{\Sigma}_n) = \Sigma$. La loi des grands nombres appliquée aux $d \times d$ coefficients donne

$$\widehat{\Sigma}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \Sigma.$$

On souhaite étudier le comportement de la matrice aléatoire $\widehat{\Sigma}_n$ lorsque la dimension des données $d = d_n$ dépend de n et tend vers ∞ . Cela nous conduit au modèle simplifié suivant : on se donne une famille $(Y_{ij})_{i,j \geq 1}$ de v.a.r. i.i.d. de moyenne 0 et de variance 1, et pour tout entier $n \geq 1$, on considère la matrice aléatoire $d_n \times d_n$ symétrique

$$\frac{1}{n} Y Y^\top$$

où $Y = (Y_{ij})_{1 \leq i \leq d_n, 1 \leq j \leq n}$. Si $n \mapsto d_n$ est constante et égale à d alors $\frac{1}{n} Y Y^\top$ converge p.s. vers $\Sigma = I_d$. Il est naturel de chercher à comprendre le comportement de la matrice $\frac{1}{n} Y Y^\top$ lorsque d_n dépend de n , par exemple en étudiant son spectre. L'analyse de la matrice symétrique $Y Y^\top$ est rendue difficile par le fait que ses coefficients sont dépendants, contrairement à ceux des matrices $Y + Y^\top$ du théorème 1.3 de Wigner.

On dispose du théorème 3.1 ci-dessous, qui se démontre comme le théorème de Wigner, bien que la mise en œuvre soit plus lourde même dans le cas carré $d_n = n$. Il a été obtenu à la fin des années 1960 par Marchenko et Pastur. Nous renvoyons à [BS] et à [PS] pour la preuve.

Théorème 3.1 (Marchenko-Pastur rectangulaire). *Si $\Sigma = I_{d_n}$ et si $\rho := \lim_{n \rightarrow \infty} \frac{d_n}{n}$ est > 0 et fini, alors, presque sûrement, pour toute fonction continue et bornée $f : \mathbb{R} \rightarrow \mathbb{R}$,*

$$\int f d\mu_{\frac{1}{n} Y Y^\top} \xrightarrow[n \rightarrow \infty]{} \int f d\mu$$

où μ est la loi de Marchenko-Pastur

$$q\delta_0 + \frac{1}{\rho 2\pi x} \sqrt{(b-x)(x-a)} \mathbf{1}_{[a,b]}(x) dx.$$

avec

$$q = \max(0, (1 - \rho^{-1})) \quad \text{et} \quad a = (1 - \sqrt{\rho})^2 \quad \text{et} \quad b = (1 + \sqrt{\rho})^2.$$

De plus, les moments de μ sont donnés pour tout $r \in \mathbb{N}^*$ par

$$\int x^r d\mu(x) = \sum_{k=0}^{r-1} \frac{\rho^k}{k+1} \binom{r}{k} \binom{r-1}{k}.$$

En particulier, la loi μ a pour moyenne 1 et variance ρ .

La transformée de Cauchy-Stieltjes de la loi μ du théorème 3.1 est donnée par

$$S_\mu(z) = \frac{1 - \rho - z + \sqrt{(z - 1 - \rho)^2 - 4\rho}}{2\rho z}$$

pour tout $z \in \mathbb{C}_+$. Il s'agit de l'unique solution à partie imaginaire positive de l'équation du second degré en s donnée par $\rho z s^2 + (z + \rho - 1)s + 1 = 0$, qui s'écrit également

$$s = \frac{1}{1 - z - \rho - \rho z s}.$$

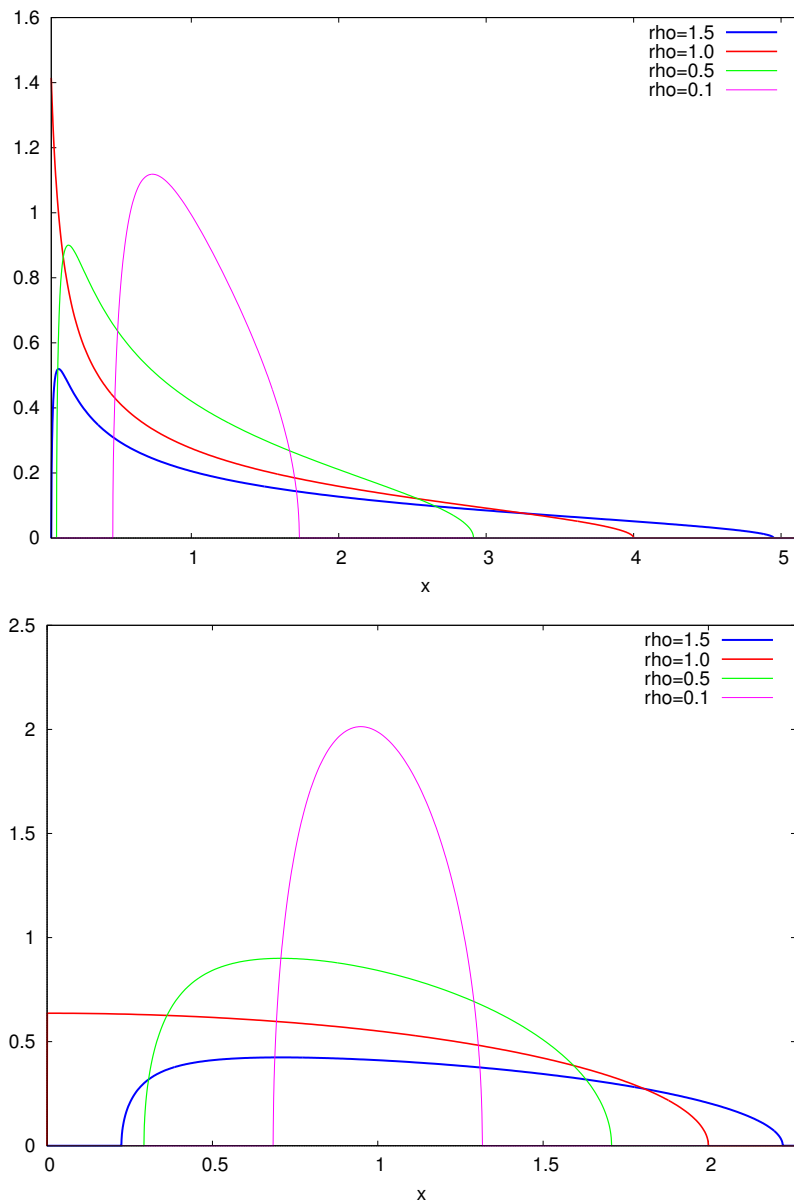


Figure 6 – Densité de Marchenko-Pastur pour différentes valeurs de ρ . Les graphiques du haut correspondent au spectre de $\frac{1}{n}YY^T$ et ceux du bas au spectre de $\sqrt{\frac{1}{n}YY^T}$. Ce graphique – obtenu avec wxMaxima (Debian GNU/Linux) – est tiré de [CG+].

La loi μ dans le théorème 3.1 est un mélange entre une masse de Dirac en 0 et une loi à densité. L'atome en 0 disparaît lorsque $\rho \leq 1$. Le théorème 3.1 affirme que p.s., pour tout $x \in \mathbb{R}$, avec $x \neq 0$ si $\rho > 1$, en notant $I =]-\infty, x]$,

$$\lim_{n \rightarrow \infty} \mu_n(I) = \mu(I).$$

Lorsque $\rho = 1$, la matrice $\frac{1}{n}YY^\top$ est en quelque sorte asymptotiquement carrée. Dans ce cas, $a = 0$ et $b = 4$. C'est le cas par exemple si $d_n = n$ pour tout $n \geq 1$. Dans ce cas, la loi μ est l'image de la loi du quart de cercle sur $[0, 2]$ par l'application $x \mapsto x^2$. On retrouve ainsi le théorème 1.2 par changement de variable.

Lorsque les coefficients de Y sont i.i.d. gaussiens standard sur \mathbb{C} (respectivement sur \mathbb{R}), la loi de YY^* est appelée loi de Wishart complexe (respectivement réelle), et constitue une sorte de loi du χ^2 matriciel. La loi des valeurs propres est alors appelée Laguerre Unitary Ensemble (respectivement Laguerre Orthogonal Ensemble).

3.4 Quelques mots sur le théorème de Girko

Les théorèmes 1.2 et 1.3 de Marchenko-Pastur et de Wigner concernent des modèles de matrices aléatoires hermitiennes. Les matrices hermitiennes sont diagonalisables en base orthonormée, car elles appartiennent aux matrices normales⁹. Le modèle le plus agréable de matrices aléatoires normales non-hermitiennes est sans doute celui fourni par la loi uniforme (ou mesure de Haar normalisée) sur le groupe unitaire \mathbb{U}_n . De telles matrices sont appelée *Haar unitaire*. Cette loi est unitairement invariante car la mesure de Haar est invariante par translation. La densité des valeurs propres d'une matrice Haar unitaire se calcule et fait apparaître un déterminant de Vandermonde, tout comme pour le GUE. La mesure empirique des valeurs propres converge vers la loi uniforme sur le cercle unité quand la dimension tend vers l'infini.

Des trois théorèmes 1.1-1.2-1.3, le théorème 1.1 de Girko semble être le plus simple. C'est pourtant le plus difficile à démontrer, car la méthode des moments et la méthode de la résolvante sont mal adaptées aux matrices non normales. Il a été imaginé par Girko dans les années 1980, en s'inspirant du résultat connu pour le cas gaussien étudié par Mehta et Ginibre¹⁰. Notons que si les coefficients de M ont une densité, alors $\mathbb{P}(MM^* = M^*M) = 0$ et donc presque sûrement M n'est pas normale !

La preuve complète du théorème 1.1 a été obtenue il y a quelques années par Tao et Vu, après cinquante ans de résultats partiels par plusieurs vagues de mathématiciens. L'approche fait appel au potentiel logarithmique, lié au laplacien. Le potentiel logarithmique au point $z \in \mathbb{C}$ d'une loi μ sur \mathbb{C} à support compact est défini par

$$U_\mu(z) := \int \ln \frac{1}{|z - \lambda|} d\mu(\lambda).$$

Le potentiel logarithmique U_μ permet de reconstituer μ car $\Delta U_\mu = -2\pi\mu$, au sens où

$$\int_{\mathbb{R}^2} U(z)\varphi(z) dx dy = -2\pi \int \Delta\varphi(\lambda) d\mu(\lambda)$$

pour tout $\varphi : \mathbb{C} \rightarrow \mathbb{R}$ C^∞ et à support compact, avec $\Delta := \partial_x^2 + \partial_y^2$ (laplacien). Cette formule peut être vue comme une conséquence de la formule de Cauchy-Pompeiu,

9. Ici $A \in \mathcal{M}_n(\mathbb{C})$ est normale quand $AA^* = A^*A$, et cela n'a rien à voir avec la loi normale !

10. L'analogie du GUE, l'ensemble de Ginibre complexe, est plus agréable à étudier que le GUE !

et dit essentiellement que $-\frac{1}{2\pi} \ln |\cdot|$ est la solution fondamentale de l'équation de la chaleur sur \mathbb{R}^2 . Le point de départ de Girko pour établir le théorème 1.1 est la formule d'hermitisation suivante, valable pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ et tout $z \in \mathbb{C}$:

$$\begin{aligned} -U_{\mu_A}(z) &= \int \ln |z - \lambda| d\mu_A(\lambda) \\ &= \frac{1}{n} \ln |\det(A - zI)| \\ &= \frac{1}{n} \ln \det(\sqrt{(A - zI)(A - zI)^*}) \\ &= \int_0^\infty \ln(t) d\mu_{\sqrt{(A - zI)(A - zI)^*}}(t). \end{aligned}$$

Elle ramène la preuve du théorème 1.1 à établir tout d'abord que presque sûrement, pour tout $z \in \mathbb{C}$, la loi $\mu_{\sqrt{(n^{-1/2}M - zI)(n^{-1/2}M - zI)^*}}$ converge en loi vers une loi ν_z sur \mathbb{R}_+ telle que $\frac{1}{2\pi} \Delta \int_0^\infty \ln(t) d\nu_z(t)$ est la loi uniforme sur le disque. Il s'agit d'étendre le théorème 1.2 à la matrice translatée $n^{-1/2}M - zI$. Le second ingrédient nécessaire à la preuve du théorème 1.1 consiste à contourner la dégénérescence de la fonction \ln en $t = \infty$ et surtout en $t = 0$, en contrôlant le bord du spectre de la matrice hermitienne semi-définie positive $(n^{-1/2}M - zI)(n^{-1/2}M - zI)^*$. Nous renvoyons par exemple à [BC] pour plus d'information. Ce dernier problème est relié à la conjecture suivante : si A est une matrice aléatoire $n \times n$ à coefficients ± 1 i.i.d. de loi de Rademacher centrée, c'est-à-dire que $\mathbb{P}(A_{ij} = 1) = \mathbb{P}(A_{ij} = -1) = 1/2$ pour tous i, j , alors

$$\mathbb{P}(\det(A) = 0) = \left(\frac{1}{2} + o_{n \rightarrow \infty}(1) \right)^n.$$

La probabilité d'égalité de deux lignes ou de deux colonnes dans A vaut 2^{-n} . L'invertibilité des matrices aléatoires est un thème actuel, situé entre analyse géométrique de grande dimension [EK, Ch. 5], combinatoire additive [TV], et théorie des probabilités.

A Quelques outils

Cette section rassemble quelques outils à la fois beaux et utiles. Nous commençons par deux théorèmes dont la preuve se trouve par exemple dans [HJ, Bha].

Théorème A.1 (Formules variationnelles min-max de Courant-Fischer). *Soit $A \in \mathcal{M}_n(\mathbb{C})$ hermitienne de valeurs propres $\lambda_1(A) \geq \dots \geq \lambda_n(A)$. Alors pour tout $1 \leq k \leq n$, en notant \mathcal{G}_k l'ensemble des sous-espaces vectoriels de \mathbb{C}^n de dimension k , on a*

$$\lambda_k(A) = \max_{V \in \mathcal{G}_k} \min_{\substack{x \in V \\ \|x\|_2=1}} \langle Ax, x \rangle = \min_{V \in \mathcal{G}_{n-k+1}} \max_{\substack{x \in V \\ \|x\|_2=1}} \langle Ax, x \rangle.$$

On en déduit un entrelacement pour le spectre des perturbations additives à rang contrôlé : si $A, B \in \mathcal{M}_n(\mathbb{C})$ sont hermitiennes de valeurs propres $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ et $\lambda_1(B) \geq \dots \geq \lambda_n(B)$ et si $r := \text{rang}(B - A)$ alors pour tout $1 \leq k \leq n$,

$$\lambda_{k+r}(A) \leq \lambda_k(B) \leq \lambda_{k-r}(A)$$

avec la convention $\lambda_k(A) = +\infty$ si $k < 1$ et $\lambda_k(A) = -\infty$ si $k > n$. Il en découle que si

$$F_A(t) := \frac{\text{card}\{1 \leq k \leq n : \lambda_k(A) \leq t\}}{n} \quad \text{et} \quad F_B(t) := \frac{\text{card}\{1 \leq k \leq n : \lambda_k(A) \leq t\}}{n}$$

sont les fonctions de répartition des mesures de comptage

$$\mu_A := \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k(A)} \quad \text{et} \quad \mu_B := \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k(B)}$$

alors

$$\|F_A - F_B\|_\infty := \sup_{t \in \mathbb{R}} |F_A(t) - F_B(t)| \leq \frac{\text{rang}(A - B)}{n}.$$

Le membre de gauche est la distance de Kolmogorov-Smirnov entre les lois de probabilités discrètes μ_A et μ_B . Voici un exemple d'application : si H est une matrice aléatoire hermitienne telle que $(H_{i,j})_{1 \leq i,j \leq n}$ sont de même moyenne m alors on a $\text{rang}(\mathbb{E}(H)) = \text{rang}(m1_n) \leq 1$ et donc $\mu_{H - \mathbb{E}(H)}$ et μ_H ont les mêmes limites en loi !

Théorème A.2 (Inégalité de Hoffman-Wielandt). *Si $A, B \in \mathcal{M}_n(\mathbb{C})$ sont hermitiennes de spectres respectifs $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ et $\lambda_1(B) \geq \dots \geq \lambda_n(B)$ alors*

$$\sum_{k=1}^n (\lambda_k(A) - \lambda_k(B))^2 \leq \sum_{i=1}^n \sum_{j=1}^n |A_{ij} - B_{ij}|^2.$$

Notons \mathcal{P}_2 l'ensemble des lois sur \mathbb{R} possédant un moment d'ordre 2 fini. La distance de couplage (ou de Wasserstein) d sur \mathcal{P}_2 est définie pour tous $\nu_1, \nu_2 \in \mathcal{P}_2$ par

$$d(\nu_1, \nu_2)^2 := \inf_{\substack{(X_1, X_2) \\ X_1 \sim \nu_1, X_2 \sim \nu_2}} \mathbb{E}(|X_1 - X_2|^2)$$

où l'infimum porte sur l'ensemble des couples de variables aléatoires de lois marginales ν_1 et ν_2 . La convergence pour d entraîne la convergence en loi (ainsi que la convergence des deux premiers moments). Dans le cas de loi discrètes de la forme $\nu_1 = \frac{1}{n} \sum_{k=1}^n \delta_{a_k}$ et $\nu_2 = \frac{1}{n} \sum_{k=1}^n \delta_{b_k}$ avec $a_1 \leq \dots \leq a_n$ et $b_1 \leq \dots \leq b_n$, on trouve

$$d(\nu_1, \nu_2)^2 = \frac{1}{n} \sum_{k=1}^n (a_k - b_k)^2.$$

Par ailleurs, rappelons que $\mathcal{M}_n(\mathbb{C})$ est un espace de Hilbert pour le produit scalaire $\langle A, B \rangle := \text{Tr}(AB^*)$. La norme associée est appelée norme de Schur, de Frobenius, ou encore de Hilbert-Schmidt : $\|A\|_{\text{HS}}^2 = \text{Tr}(AA^*) = \sum_{i=1}^n \sum_{j=1}^n |A_{ij}|^2 = \sum_{i=1}^n \lambda_i(A)^2$. L'inégalité de Hoffman-Wielandt peut s'écrire à présent

$$d(\mu_A, \mu_B)^2 \leq \frac{\|A - B\|_{\text{HS}}^2}{n}.$$

Voici un exemple d'application : si H est une matrice hermitienne aléatoire à valeurs dans $\mathcal{M}_n(\mathbb{C})$ telle que $(H_{ii})_{1 \leq i \leq j \leq n}$ sont i.i.d. de carré intégrable, alors pour tout réel $C > 0$, en désignant par H_C la matrice tronquée définie par $(H_C)_{ij} := H_{ij} \mathbf{1}_{\{|H_{ij}| \leq C\}}$, et en utilisant l'inégalité ci-dessus et la loi forte des grands nombres, on a

$$d(\mu_H, \mu_{H_C})^2 \leq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |H_{ij}|^2 \mathbf{1}_{\{|H_{ij}| > C\}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}(|H_{12}|^2 \mathbf{1}_{\{|H_{12}| > C\}}).$$

Par convergence dominée ou monotone, le membre de droite peut être rendu arbitrairement proche de 0 en choisissant C assez grand. Cela permet de se ramener, dans la

preuve du théorème de Wigner, au cas où les coefficients de H sont bornés. La même méthode permet d'établir que la diagonale de H ne joue pas de rôle dans le théorème de Wigner. On peut en particulier la supposer nulle, ou de variance arbitraire.

Le théorème suivant se trouve par exemple dans [MD].

Théorème A.3 (Inégalité de concentration de Azuma-Hoeffding). *Si $G(X_1, \dots, X_n)$ est une variable aléatoire intégrable fonction de vecteurs aléatoires X_1, \dots, X_n indépendants (pas forcément de même dimension), alors pour tout $r \geq 0$,*

$$\mathbb{P}(|G(X_1, \dots, X_n) - \mathbb{E}(G(X_1, \dots, X_n))| \geq r) \leq 2 \exp\left(-\frac{r^2}{2(c_1^2 + \dots + c_n^2)}\right)$$

où $c_k := \sup_{(x, x') \in \mathcal{D}_k} |G(x) - G(x')|$ et $\mathcal{D}_k := \{(x, x') : x_i = x'_i \text{ si } i \neq k\}$.

Cette inégalité, de même nature que celle de Bienaymé-Tchebychev¹¹, affirme que la variable aléatoire $G(X_1, \dots, X_n)$ est concentrée autour de sa moyenne. Donnons un exemple d'application pour les distributions spectrales de matrices aléatoires. Si f est une fonction de classe \mathcal{C}^1 et à support compact, alors pour toutes matrices hermitiennes $A, B \in \mathcal{M}_n(\mathbb{C})$, en notant $r := \text{rang}(A - B)$, nous savons déjà, via le théorème A.1, que $\|F_A - F_B\|_\infty \leq r/n$, et donc, par intégration par parties,

$$\left| \int f d\mu_A - \int f d\mu_B \right| = \left| \int_{-\infty}^{+\infty} f'(t)(F_A(t) - F_B(t)) dt \right| \leq \frac{r}{n} \int_{-\infty}^{+\infty} |f'(t)| dt := \frac{r}{n} \|f'\|_1.$$

On note $A(x_1, \dots, x_n) \in \mathcal{M}_n(\mathbb{C})$ la matrice hermitienne dont les lignes du triangle supérieur sont x_1, \dots, x_n (ce sont des vecteurs de $\mathbb{C}^n, \mathbb{C}^{n-1}, \dots, \mathbb{C}$), et on pose

$$G(x_1, \dots, x_n) = \int f d\mu_{A(x_1, \dots, x_n)}.$$

On observe que

$$\text{rang}\{A(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) - A(x_1, \dots, x_{k-1}, x'_k, x_{k+1}, \dots, x_n)\} \leq 2$$

et donc

$$|G(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) - G(x_1, \dots, x_{k-1}, x'_k, x_{k+1}, \dots, x_n)| \leq \frac{2}{n} \|f'\|_1.$$

À présent, si H est une matrice hermitienne aléatoire à valeurs dans $\mathcal{M}_n(\mathbb{C})$ telle que les lignes de son triangle supérieur $(H_{ij})_{1 \leq i \leq j \leq n}$ sont indépendantes (ce sont des vecteurs aléatoires de $\mathbb{C}^n, \mathbb{C}^{n-1}, \dots, \mathbb{C}$), alors, grâce à l'inégalité de Azuma-Hoeffding, pour toute fonction f de classe \mathcal{C}^1 et à support compact, et pour tout réel $r \geq 0$,

$$\mathbb{P}\left(\left| \int f d\mu_H - \mathbb{E} \int f d\mu_H \right| \geq r\right) \leq 2 \exp\left(-\frac{nr^2}{8\|f'\|_1^2}\right).$$

Cette inégalité exprime le fait que la distribution spectrale empirique est très concentrée autour de son espérance. C'est pour cette raison que les fluctuations sont si faibles dans les simulations, dès que la dimension dépasse quelques dizaines. La variable aléatoire $\int f d\mu_H - \mathbb{E} \int f d\mu_H$ converge en probabilité vers 0. Pour tirer parti de

11. Si G est une v.a.r. alors $\mathbb{P}(|G - \mathbb{E}(G)| \geq r) \leq \text{Var}(G)/r^2$ pour tout $r > 0$.

la rapidité de convergence, on pose par exemple $r = n^{-1/2+\varepsilon}$ avec $\varepsilon > 0$ petit fixé, de sorte que la borne exponentielle dans le membre de droite de l'inégalité soit sommable en n . Le lemme de Borel-Cantelli¹² entraîne alors que la convergence de la variable aléatoire vers 0 a lieu presque sûrement. Cela permet de ramener l'étude de la convergence en loi presque sûre de μ_H à celle de la convergence en loi de $\mathbb{E}\mu_H$.

Références

- [ABF] G. Akemann, J. Baik et P. Di Francesco (éds.) – *The Oxford handbook of random matrix theory*, Oxford University Press, Oxford, 2011.
- [AGZ] G. W. Anderson, A. Guionnet et O. Zeitouni – *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.
- [BC] C. Bordenave et D. Chafaï – « Around the circular law », *Probab. Surv.* **9** (2012), p. 1–89.
- [Bha] R. Bhatia – *Matrix analysis*, Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997.
- [Bi1] P. Biane – « Actes des journées X-UPS 2002 », ch. La fonction Zêta de Riemann et les probabilités, p. 165–193, Ed. Éc. Polytech., 2003.
- [Bi2] —, « Probabilités libres et matrices aléatoires », Image des maths, 2004.
- [BS] Z. Bai et J. W. Silverstein – *Spectral analysis of large dimensional random matrices*, second éd., Springer Series in Statistics, Springer, New York, 2010.
- [CD] R. Couillet et M. Debbah – *Random matrix methods for wireless communications*, Cambridge University Press, Cambridge, 2011.
- [CG+] D. Chafaï, O. Guédon, G. Lecué, et A. Pajor, *Interactions between compressed sensing, random matrices, and high dimensional geometry*. Panoramas et Synthèses numéro 37, Société Mathématique de France, à paraître, 2012.
- [Da] S. Dallaporta – *Quelques aspects de l'étude quantitative de la fonction de comptage et des valeurs propres de matrices aléatoires*, thèse de doctorat de l'Université de Toulouse, 2012. PDF
- [De] P. A. Deift – *Orthogonal polynomials and random matrices : a Riemann-Hilbert approach*, Courant Lecture Notes in Mathematics, vol. 3, New York University Courant Institute of Mathematical Sciences, New York, 1999.
- [DG] P. Deift et D. Gioev – *Random matrix theory : invariant ensembles and universality*, Courant Lecture Notes in Mathematics, vol. 18, Courant Institute of Mathematical Sciences, New York, 2009.
- [EK] Y. C. Eldar et G. Kutyniok (éds.) – *Compressed sensing*, Cambridge University Press, Cambridge, 2012, Theory and applications.
- [ER] A. Edelman et N. R. Rao – « Random matrix theory », *Acta Numer.* **14** (2005), p. 233–297.
- [E] L. Erdős – « Universality of Wigner random matrices : a survey of recent results », *Russian Mathematical Surveys* **66** (2011), p. 507–626.
- [F] P. J. Forrester – *Log-gases and random matrices*, London Mathematical Society Monographs Series, vol. 34, Princeton University Press, Princeton, NJ, 2010.
- [G1] A. Guionnet – *Large random matrices : lectures on macroscopic asymptotics*, Lecture Notes in Mathematics, vol. 1957, Springer-Verlag, Berlin, 2009, Lectures from the 36th Probability Summer School held in Saint-Flour, 2006.
- [G2] —, « Grandes matrices aléatoires et théorèmes d'universalité (d'après Erdős, Schlein, Tao, Vu et Yau) », *Astérisque* (2011), no. 339, p. Exp. No. 1019, viii, 203–237, Séminaire Bourbaki. Vol. 2009/2010. Exposés 1012–1026.
- [HJ] R. A. Horn et C. R. Johnson – *Matrix analysis*, second éd., Cambridge University Press, Cambridge, 2013.
- [HP] F. Hiai et D. Petz – *The semicircle law, free random variables and entropy*. Mathematical Surveys and Monographs, 77. American Mathematical Society, Providence, RI, 2000. x+376 pp.

12. Si $\mathbb{E}(\sum_n \mathbf{1}_{A_n}) = \sum_n \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\sum_n \mathbf{1}_{A_n} < \infty) = 1$. Note : $\{\sum_n \mathbf{1}_{A_n} = \infty\} = \bigcap_n \bigcup_{m \geq n} A_m$.

- [L] M. Ledoux – « Differential operators and spectral distributions of invariant ensembles from the classical orthogonal polynomials. The continuous case », *Electron. J. Probab.* **9** (2004), p. no. 7, 177–208 (electronic).
- [M] M. L. Mehta – *Random matrices*, third éd., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004.
- [MD] C. McDiarmid – *On the method of bounded differences*, in *Surveys in combinatorics, 1989* (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, p. 148–188.
- [P] S. Péché – « La plus grande valeur propre de matrices de covariance empirique », *Image des maths*, 2006.
- [PS] L. Pastur et M. Shcherbina – *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs, vol. 171, American Mathematical Society, Providence, RI, 2011.
- [T1] T. Tao – *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012.
- [T2] T. Tao – *The asymptotic distribution of a single eigenvalue gap of a Wigner matrix*, prépublication 2012, disponible sur arXiv :1203.1605, et commentée sur le blog de l’auteur.
- [TV] T. Tao et V. Vu – *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, 105, Cambridge University Press, Cambridge, 2006.



Figure 7 – Timbre-poste hongrois en l’honneur de Wigner.

Remerciements. Ce texte a bénéficié de commentaires de relecture de Charles Bordenave, Yan Doumerc, Christophe Giraud, Florent Malrieu, et Pierre-André Zitt.